Event Detection through Differential Pattern Mining in Internet of Things

Authors:
Md Zakirul Alam Bhuiyan and Jie Wu

IEEE MASS 2016
The 13th IEEE International Conference on Mobile Ad hoc and Sensor Systems
Brasília, Brazil, October 10-13, 2016

Outline

- Motivation
- Existing Work
- Proposed Scheme
- Data Preparation
- Mining through DP-Tree Development
- Performance Evaluations
- Conclusions and Future Works

Motivations

- Internet of Things (IoT) have strong practical applications in many domains, e.g.,
  - Structural health monitoring (SHM) for industrial machine, aerospace, and vehicles
  - Chemical explosion, military surveillance, intrusion tracking.

  - In these applications, high quality event detection using wireless sensing in IoT is essential.

Motivations

- Wireless sensors in IoT produce a huge volume of dynamic data when deployed in these applications.
  - **It is vital to develop methodologies to mine the big data**
    - To detect event of interest in the applications
      - With low cost and high-quality

Outline

- Motivation
- **Existing Work**
- Proposed Scheme
- Data Preparation
- Mining through DP-Tree Development
- Performance Evaluations
- Conclusions and Future Works

Existing Work

- Traditional data mining schemes used to mine data in IoT
  - Frequent pattern
    - Association rules
    - Sequential pattern
  - Clustering
  - Classification

- Centralized and computationally expensive
- Focusing on transactional data
- Data patterns are straightforward

- Sensors in IoT may face difficulties in providing event information

Existing Work

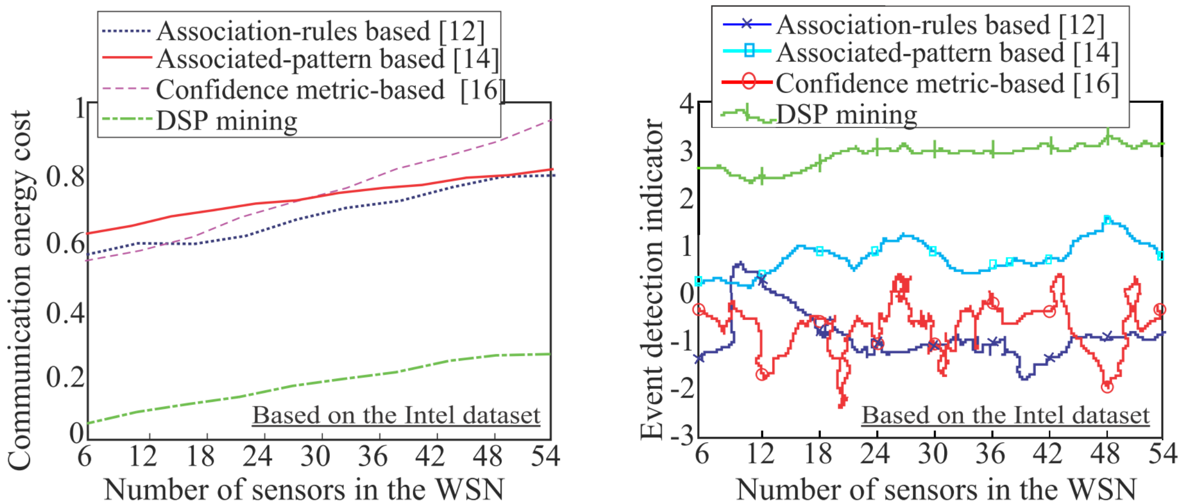
- Traditional data mining schemes used to mine data in IoT
  - **Frequent pattern**
    - Association rules
    - Sequential pattern
  - **Clustering**
  - **Classification**
- Sensors in IoT may face difficulties in providing event information

Damage, crack, explosive, fire, mobile event (via sensing signals or Wi-Fi signals)

0/1 pattern, sum, avg., max, metric → Decision-making → Meaningful decision for an event?

Existing Work

- Performance comparison: when using event indicator



(i) Communication energy cost analysis (ii) Event detection in different schemes

Outline

- Motivation
- Existing Work
- **Proposed Scheme**
- Data Preparation
- Mining through DP-Tree Development
- Performance Evaluations
- Conclusions and Future Works

Problem Definition

- Network model
  - A wireless sensor network
    - A set of energy-constrained sensors
      - Organized into CHs connecting a BS
    - A minimum communication range, sensors are allowed to share their mined information with their neighbors
  - A computation and communication cost models are given
- Target application: SHM, smart city applications

Problem Definition

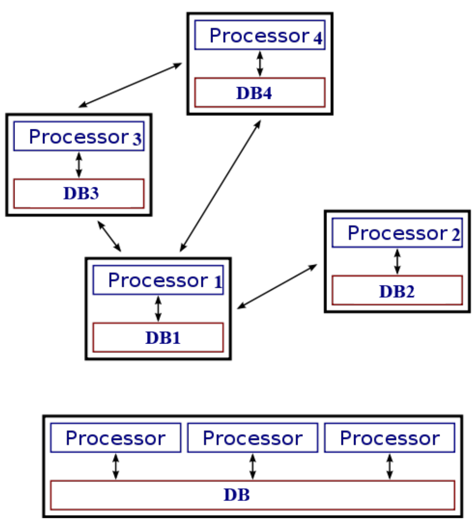
- **Find:**
  - **A pattern of sensors (that may report an event information)**
    - By mining all the acquired data of sensors in a cluster in a distributed and parallel manner such that a CH can finally decide whether an event has occurred in the area and report to the BS.
- **Objectives:**
  - **To reduce the communication cost of the wireless sensor and to provide high-quality event detection.**

Our Scheme

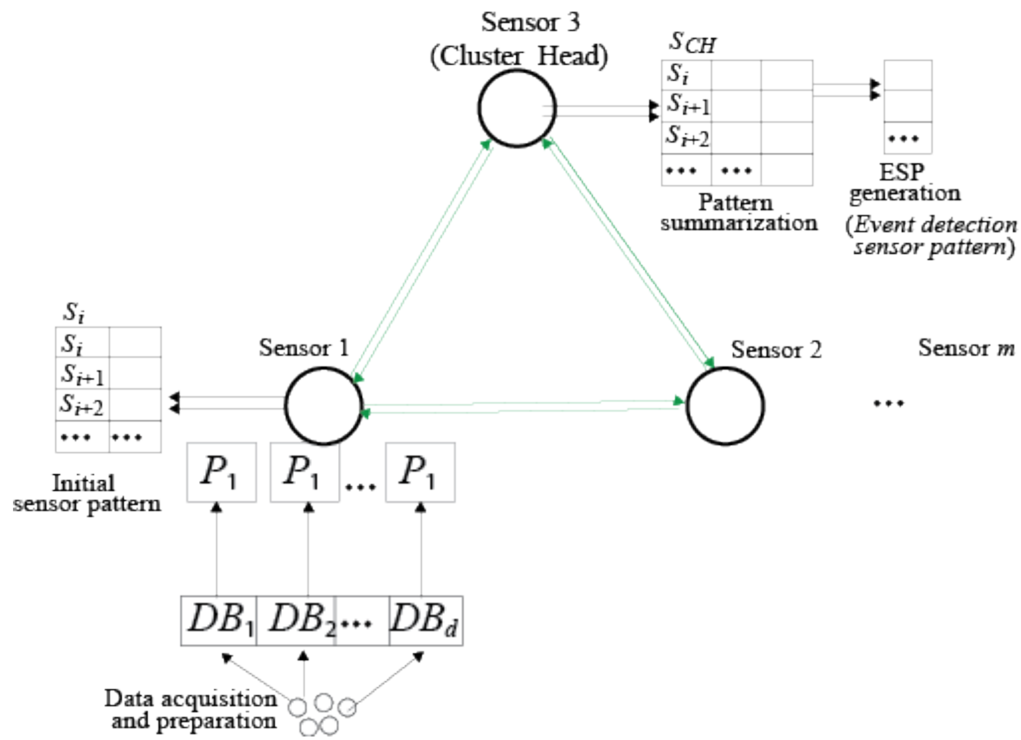
- **DPminer**
  - **A sensor data mining scheme for event detection**
    - Supports IoT applications
  - **Function in a distributed and parallel manner**
    - Data in a partitioned database processed in a distributed and parallel manner by one or more sensor processors
    - Finding differential information between sensor databases and extracting a pattern of sensors having event information
      - DSP: Differential event-sensitive sensor pattern

Our Scheme

- DPminer
  - **A sensor data mining scheme for event detection**
    - Supports IoT applications
  - **Function in a distributed and parallel manner**

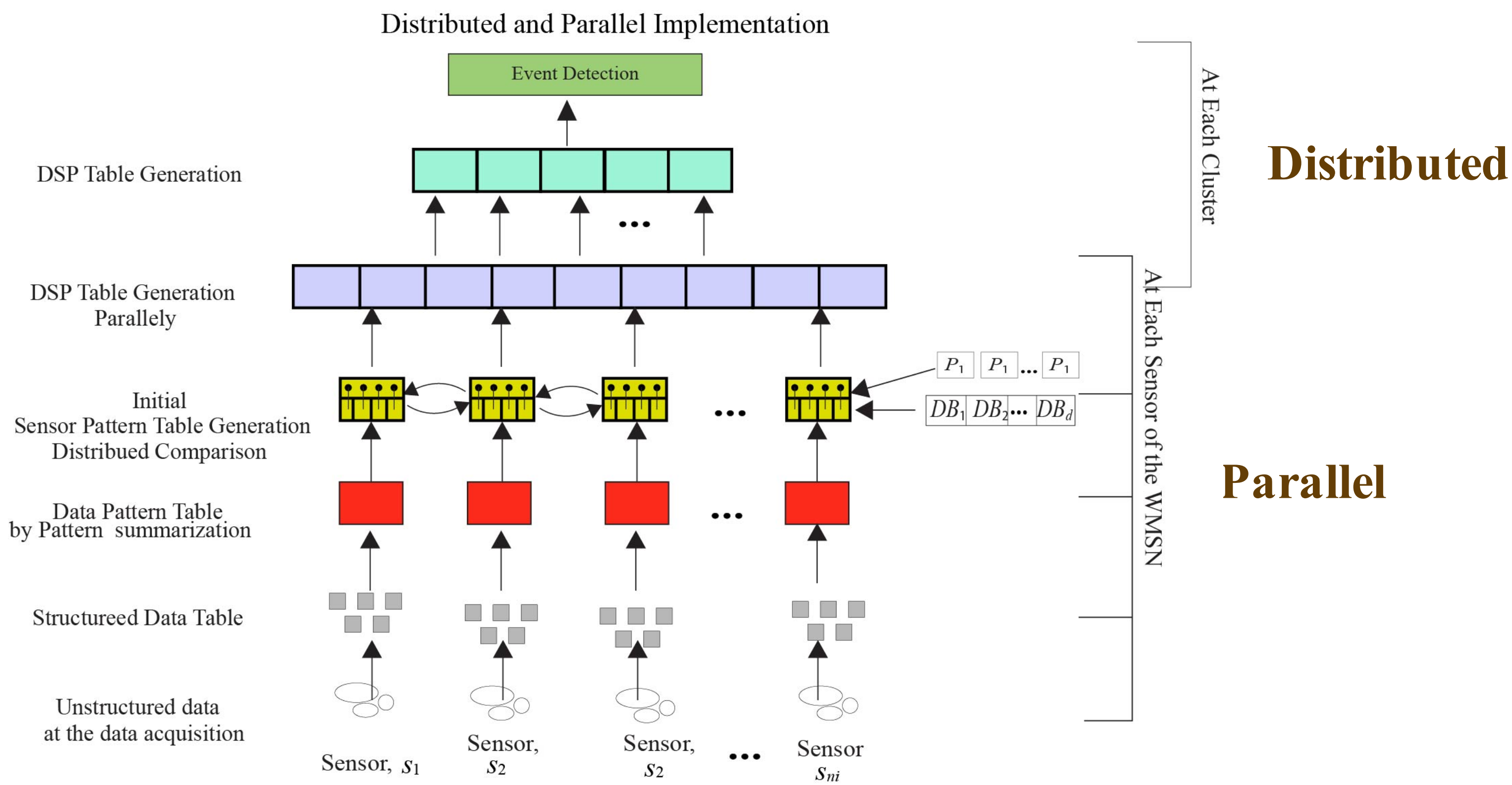


Using the similar concept



Our Scheme

- DSP: Differential Pattern Mining



Outline

- Motivation
- Existing Work
- Proposed Scheme
- **Data Preparation**
- Mining through DP-Tree Development
- Performance Evaluations
- Conclusions and Future Works

Data Preparation: The 1st Stage Mining

- Proportion test
  - **Acquired raw data**
    - Unstructured, noisy, incomplete, out of range
  - **Maintain two databases**
    - Control database
      - Containing a simplified dataset– a set of ranges and a set of tuples with different frequencies and values that can be defined by the healthy data (when there is no event in an application).
    - Case database
      - Stored data collected at a specific period
  - **Proportion test is used to check: whether the data in the case is within a given range or not through a comparison between the Control and the Case**

Data Preparation: The 1st Stage Mining

- Proportion test

$$z = \frac{p_{case} - p_{control}}{\sqrt{p(1-p)(\frac{1}{\pi_{case}} + \frac{1}{\pi_{control}})}}$$

- $H_{in} : \pi_{Case} = \pi_{Control}$ vs $H_{out} : \pi_{Case} \neq \pi_{Control}$
  - H_{in} and H_{out} denote the frequencies and values that are ‘in’ the range and ‘out’ of the range, respectively in between Cases and Controls dataset.
- Under the null hypothesis of no difference in values, the square of the statistic z^2 follows the **Pearson’s chi-squared test**

Data Preparation: The 1st Stage Mining

- Data Summarization
  - Values, frequencies

Sensor s_1

(i) Structured segmened values at t_1

d1	0.05685	0.18652	0.12451	0.21546	0.06592	0.18652	…
d2	0.12596	0.01256	0.12981	0.26451	0.29865	0.08289	…
d3	0.01652	0.16029	0.17045	0.01421	0.02429	0.19077	…
…	…	…	…	…	…	…	…

(ii) Data ranging

A=0.05	B=0.15	C=0.25	D=0.35	E=0.45	F=0.55	G=0.65	H=0.75	I=0.85	J=0.95
0.00 0.09	0.10 0.19	0.21 0.29	0.31 0.39	0.41 0.49	0.51 0.59	0.61 0.69	0.71 0.79	0.81 0.89	0.91 1.00
low to high event intensity									
low event intensity			med event intensity			high event intensity			

(iii) Labeling Values

H_1	A	B	B	C	A	B	C	D	B	A	B	D	D	C	B	B
H_2	B	A	B	C	C	A	B	D	A	E	C	A	B	A	B	A
H_3	A	B	B	A	C	B	B	A	F	D	A	A	B	B	B	E
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

(iv) Calculating Frequencies

A=3	B=7	C=3	D=3		
A=6	B=5	C=3	D=1	E=1	
A=5	B=7	C=1	D=1	E=1	F=1

(v) Total average values, T_{mv}= 8.5
Rate of Frequencies, F_{t_1}={3,3,3,3,2,1}

(vi) Summarization

Total values	TA=14	TB=19	TC=7	TD=5	TE=2	TF=1
Rate of Frequencies	FA=3	FB=3	FC=3	FD=3	FE=2	FF=1

Outline

- Motivation
- Existing Work
- Proposed Scheme
- Data Preparation
- **Mining through DP-Tree Development**
- Performance Evaluations
- Conclusions and Future Works

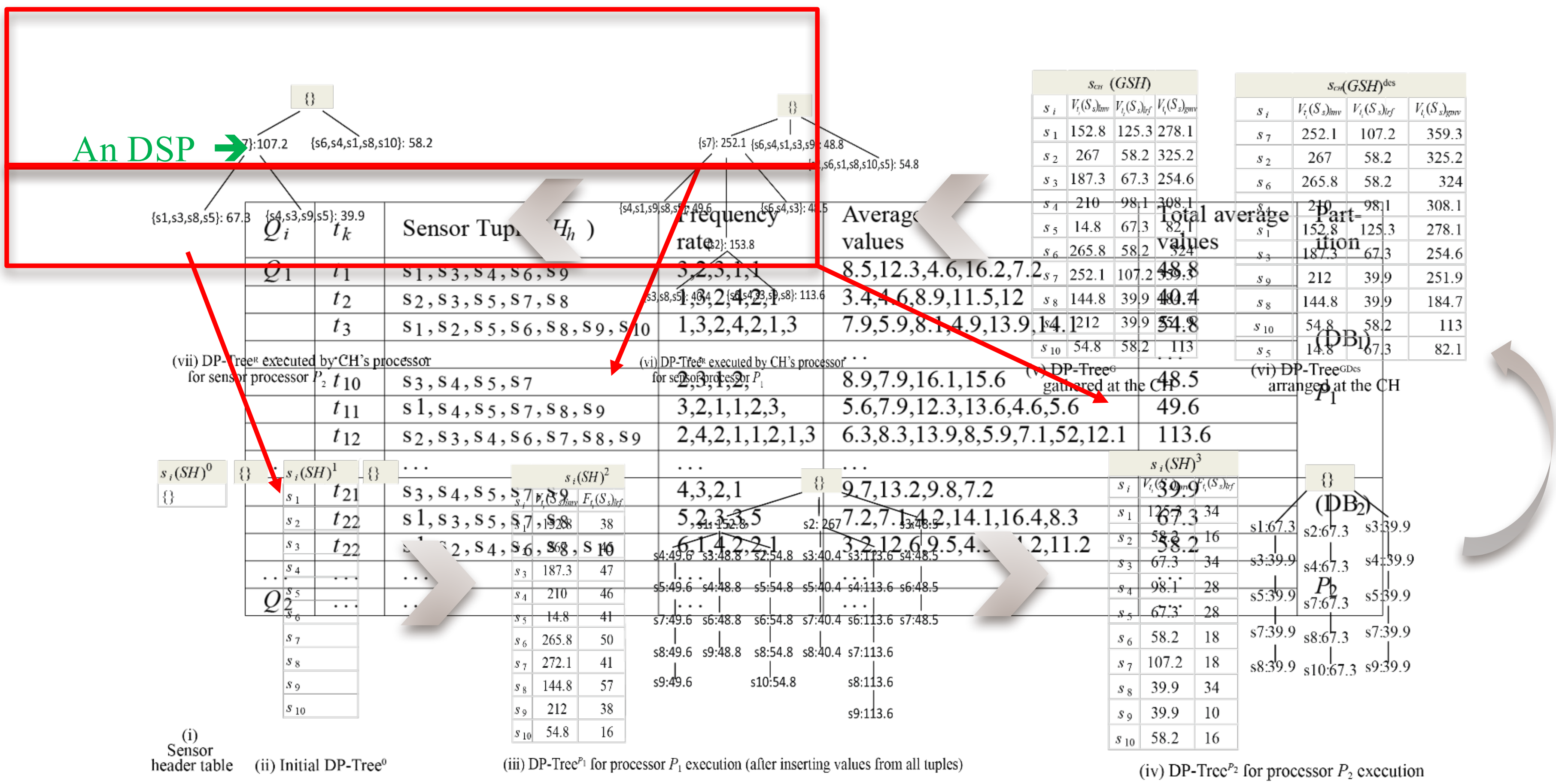
Mining through DP-Tree Development: The 2nd Stage Mining

- Data Pattern Table

Q_i	t_k	Sensor Tuple (H_h)	Frequency rate	Average values	Total average values	Part-ition
Q_1	t_1	s_1,s_3,s_4,s_6,s_9	3,2,3,1,1	8.5,12.3,4.6,16.2,7.2	48.8	(DB_1) P_1
	t_2	s_2,s_3,s_5,s_7,s_8	1,3,2,4,2,1	3.4,4.6,8.9,11.5,12	40.4	
	t_3	$s_1,s_2,s_5,s_6,s_8,s_9,s_{10}$	1,3,2,4,2,1,3	7.9,5.9,8.1,4.9,13.9,14.1	54.8	
...	...	...	...	...	...	
	t_{10}	s_3,s_4,s_5,s_7	2,3,1,2,	8.9,7.9,16.1,15.6	48.5	
	t_{11}	s_1,s_4,s_5,s_7,s_8,s_9	3,2,1,1,2,3,	5.6,7.9,12.3,13.6,4.6,5.6	49.6	
	t_{12}	$s_2,s_3,s_4,s_6,s_7,s_8,s_9$	2,4,2,1,1,2,1,3	6.3,8.3,13.9,8,5.9,7.1,52,12.1	113.6	
...	...	...	...	...	...	
	t_{21}	s_3,s_4,s_5,s_7,s_9	4,3,2,1	9.7,13.2,9.8,7.2	39.9	(DB_2) P_2
	t_{22}	s_1,s_3,s_5,s_7,s_8	5,2,3,3,5	7.2,7.1,4.2,14.1,16.4,8.3	67.3	
	t_{22}	$s_1,s_2,s_4,s_6,s_8,s_{10}$	6,1,4,2,2,1	3.2,12.6,9.5,4.5,14.2,11.2	58.2	
...	...	...	...	...	...	
Q_2	...	...	...	...	...	

Mining through DP-Tree Development: The 1st Stage Mining

- DSP Computation



Outline

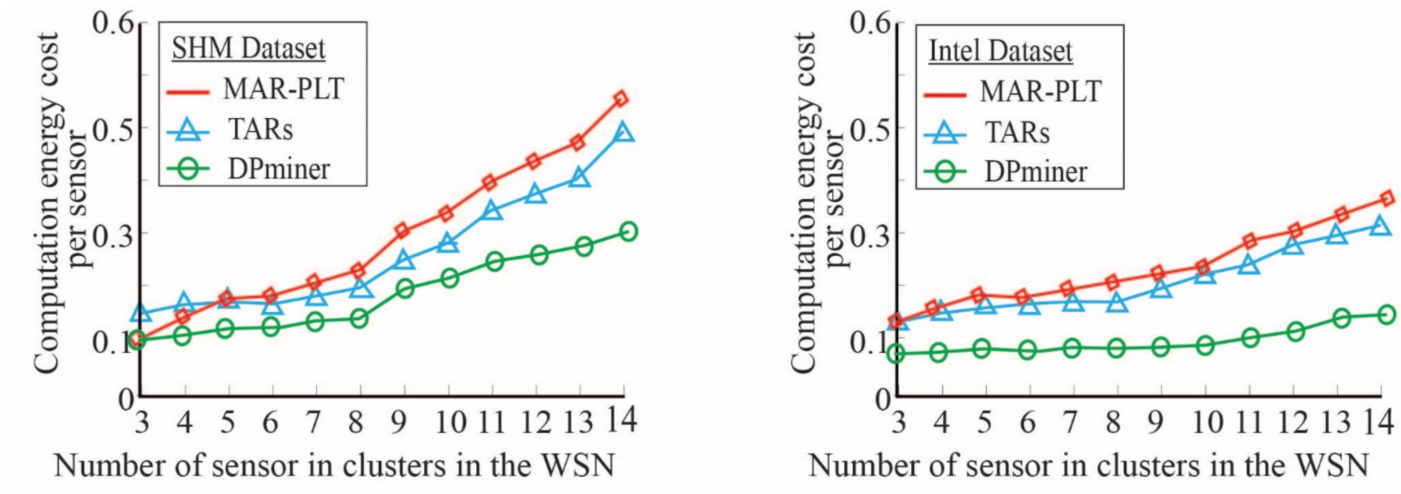
- Motivation
- Existing Work
- Proposed Scheme
- Data Preparation
- Mining through DP-Tree Development
- **Performance Evaluations**
- Conclusions and Future Works

Performance Evaluations

- For the DSP generation,
  - **Two sets of big dataset**
    - SHM data set: real data of 800 sensors collected from GNTVT
    - 54 sensors' dataset offered by Intel Berkeley Research Lab
  - **Each DB of a sensor is distributed among the processor of sensor nodes and the processors in the node has complete access to its portion of the database**
  - **Observation**
    - Computation cost, communication cost, and meaningful damage event detection information extraction
  - **A portion of data at some sensors are modified in order to provide a presence of event**

Results

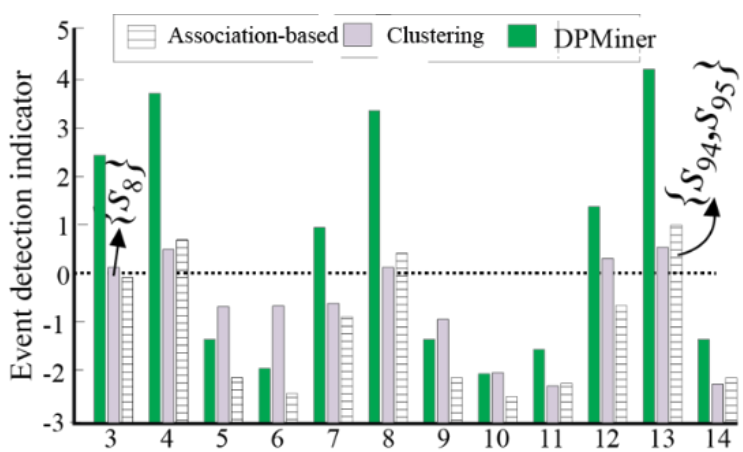
- Computation energy cost



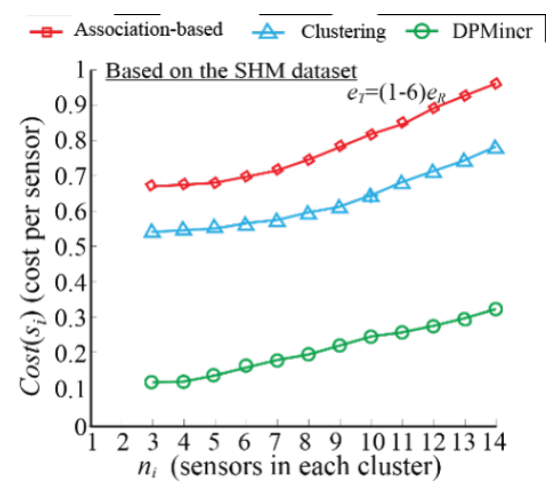
(i) Computation energy in different schemes (ii) Computation energy in different schemes

Results

- Event detection and communication energy cost



Event detection in three schemes



Communication energy cost per sensor of the WSN in different schemes

Outline

- Motivation
- Existing Work
- Proposed Scheme
- Data Preparation
- Mining through DP-Tree Development
- Performance Evaluations
- **Conclusions and Future Works**

Conclusions

- DPminer
  - A comprehensive data mining schemes for sensors in IoT
  - It works in a distributed and parallel manner and is able to extract a pattern of sensors having event information.
  - Feature: provide important values as outputs (rather than "0/1" binary decision).

- Future Work
  - Applying the differential sensor mining technique with a machine-learning approach
  - Multimedia application, biomedical application

Author Contacts:

- Md Zakirul Alam Bhuiyan
  - Email: mbhuiyan3@fordham.edu
  - https://sites.google.com/site/zakirulalam/

- Jie Wu
  - Email: jiewu@temple.edu
  - http://cis-linux1.temple.edu/~jiewu/